

Distributed Linear Precoding and User Selection in Coordinated Multicell Systems

Eduardo Castañeda, Adão Silva, Ramiro Samano-Robles, and Atílio Gameiro

Abstract

In this manuscript we tackle the problem of semi-distributed user selection with distributed linear precoding for sum rate maximization in multiuser multicell systems. A set of adjacent base stations (BS) form a cluster in order to perform coordinated transmission to cell-edge users, and coordination is carried out through a central processing unit (CU). However, the message exchange between BSs and the CU is limited to scheduling control signaling and no user data or channel state information (CSI) exchange is allowed. In the considered multicell coordinated approach, each BS has its own set of cell-edge users and transmits only to one intended user while interference to non-intended users at other BSs is suppressed by signal steering (precoding). We use two distributed linear precoding schemes, Distributed Zero Forcing (DZF) and Distributed Virtual Signal-to-Interference-plus-Noise Ratio (DVSINR). Considering multiple users per cell and the backhaul limitations, the BSs rely on local CSI to solve the user selection problem. First we investigate how the signal-to-noise-ratio (SNR) regime and the number of antennas at the BSs impact the effective channel gain (the magnitude of the channels after precoding) and its relationship with multiuser diversity. Considering that user selection must be based on the type of implemented precoding, we develop metrics of compatibility (estimations of the effective channel gains) that can be computed from local CSI at each BS and reported to the CU for scheduling decisions. Based on such metrics, we design user selection algorithms that can find a set of users that potentially maximizes the sum rate. Numerical results show the effectiveness of the proposed metrics and algorithms for different configurations of users and antennas at the base stations.

Index Terms

Semi-distributed user selection, coordinated downlink transmission, distributed linear precoding, cellular networks, interference channels.

I. INTRODUCTION

The performance of coordinated downlink transmission with linear precoding in multiple antenna multicell systems has been an active area of research over the last years. Recent works (e.g. [1]–[3] and references therein) have shown that cooperation and coordination between clustered base stations (BSs) improve rates, coverage, and efficiently suppress inter-cell interference (ICI) which specially benefits cell-edge users [4]. Multicell coordination involves message exchange between neighboring cells and according to the level of coordination, multicell systems have been classified in three groups [3], [5]–[7]: interference aware (IA), joint processing/transmission (JT), and coordinated beamforming (CBF). In IA there is no information exchange among BSs, each transmitter serves its own set of users, and transmission parameters are adjusted in a selfish fashion by measuring ICI [3]. In contrast, in JT systems it is assumed that channel state information (CSI) and user data are globally available, full coordination is attainable through a central processing unit (CU), and each user receives data from a group of coordinated BSs (cluster). The JT system can be interpreted as a broadcast channel [6] with distributed antennas and several radio resource management (RRM) tasks (e.g., scheduling, power control, precoding design, data queue control, etc.) extended from the single-cell systems can be applied (e.g., [1]–[4], [8]–[11]). However, such extensions must take into account backhaul rate limitations, CSI acquisition, joint transmission, and other system constraints [7].

In CBF the BSs need only data of the users in their own cells and they do not require to know the precoders and traffic of other BSs. The shared information is related to scheduling control signaling and CSI in order to mitigate spatial ICI. The BSs design precoding vectors towards the scheduled users so that the gain is two-fold: increasing the signal strength at the receivers and suppressing interference in the adjacent cells [7]. Efficient RRM schemes can be implemented under CBF using local CSI [12], [13] which relaxes the wideband backhaul and synchronization requirements [1].

Regardless the type of coordination between neighboring BSs, the inter-cluster interference problem arises if multiple clusters are taken into account, which can be dealt in two ways. The most straightforward way is to apply the principle of cellular planning with frequency reuse [1, §5]. Using different radio resources in adjacent clusters (it can be dynamically allocated) mitigates or eliminates the inter-cluster interference. A second approach to reuse radio resource among

different clusters is by means of inter-cluster coordination, where adjacent clusters implement interference mitigation techniques for the users at the edges of the clusters (e.g. [4]). For sake of simplicity and modeling tractability this work considers a single cluster network with B BSs for the single carrier case.

A. Related Works

Depending on the system utility function that is optimized, there exist different strategies to achieve optimal power allocation and precoding design assuming that global CSI is known, and that the number of antennas at the transmitters can serve all competing users (cf. [3], [10] for an in-depth survey). In the scenario where each BS serves only one user and CSI is not exchanged among BSs, the system model can be referred to as *interference channel* [6]. Recent works characterize its achievable rate region and jointly perform power allocation and precoding design (e.g., [5], [10], [14], [15]) under the assumption that the intended user of each BS has been previously selected by some procedure. However, for multiuser multicell scenarios each BS must select one user from its own pool of users before proceeding with precoding calculations. In this scenario the sum rate maximization is a complex combinatorial problem because the number of users is larger than the number of available spatial resources (antennas) and global CSI may not be available. The global performance is highly sensitive to the set of scheduled users, since the signal strength of the intended user k_b at BS b relies on local CSI (including the local channels of non-intended users at other BSs). Additionally, the multiplexing gain and ICI suppression depend upon the number of antennas at the BSs [7].

In the literature of single-cell MU-MISO systems with precoding based on Zero Forcing (ZF), the sum rates maximization problem is commonly tackled by decoupling the user selection from the power allocation and precoding design. The user selection is performed first based on the null space projection (NSP) (e.g., [16]–[18]) or an approximation of it (e.g., [19], [20]). The NSP provides an accurate measure of the *effective channel gain* (the channel magnitude after precoding), so that the user channels selected based on such metric are spatially compatible or quasi-orthogonal. For ZF precoding, this means that the users selected using the NSP can provide a close-to-optimal solution to the sum rate maximization problem in multiuser scenarios. Recent works on multicell systems have proposed extensions from single-cell user selection algorithms assuming that partial or global CSI is available at the CU (e.g., [2], [4], [6], [9], [11]). The

extensions in [2], [4] are centralized algorithms that exploit the concept of NSP to improve sum rates relying on global CSI at the scheduler. If global CSI is not available, distributed precoding and scheduling can still be implemented. For instance, LTE-Advanced standard [1], [21] considers distributed linear precoding such as signal-to-leakage-plus-noise ratio (SLNR) [6], [22] and ZF whose computation requires to know only local CSI and the set of intended users. One strategy for joint distributed precoding and scheduling is to limit the exchange of CSI such that the clustered BSs jointly select users in a sequential fashion, i.e., the first BS selects its user and broadcast its decision, then the second BS selects its user based on the decision made by the first one and so on [1]. Another approach has been introduced in [13] where users selection, precoding design, and power allocation are treated as decoupled problems but their parameters are jointly updated at the CU. Results show that distributed RRM schemes with limited message exchange between BSs can improve system performance.

B. Contributions

In the system model considered in this work, a set of adjacent BSs form a cluster and they coordinate their transmission strategies through a CU in order to serve a set of cell-edge users and mitigate ICI. The clustered BSs adopt the CBF transmission scheme where the data for an intended user is transmitted from one BS, whereas the impairments from the ICI are mitigated by coordinated precoding. Two distributed linear precoding schemes will be used: Distributed Zero Forcing (DZF) and Distributed Virtual Signal-to-Interference-plus-Noise Ratio (DVSINR derived from SLNR). It is assumed that each BS has its own set of intended users, no user data or CSI is exchanged between BSs, and the shared information between BSs and the CU is for scheduling control. In each scheduling instance the clustered BSs attempt to maximize the sum rate by selecting a set of users with particular characteristics. Optimizing the performance in the described scenario is a challenging task since global CSI is not available and the backhaul connection with the CU only supports scheduling control information. Moreover, selecting the best set of users whose channel characteristics maximize the sum rate is a combinatorial problem whose complexity grows exponentially with the number of BSs and users per cell [13].

To solve the user selection problem, and taking into account that the BSs implement either DZF or DVSINR, the key results of this work are summarized as follows.

- Initially, we discuss how the instantaneous and average effective channel gains of DZF and

DVSINR depend on the signal-to-noise-ratio (SNR) regime, the number of antennas at the BSs, and multiuser diversity. This insight of the precoder schemes is used to establish in which way local CSI must be processed at each BS. We design precoder-based metrics of user compatibility, i.e., depending on the type of precoding we propose a mapping from the local CSI to a real number. The proposed metrics are estimations of the achievable effective channel gains and operate in different system configurations based on the number of transmit antennas and BSs.

- The scheduling process must be performed at the CU using the metrics reported by the BSs. We accomplish this goal by developing an algorithm for user selection that properly combines the reported metrics. Once that a set of users has been selected, the decision is informed to the BSs and they compute either DZF or DVSINR based on the local CSI of the selected users.

- We propose a pre-selection methodology in order to reduce the number of competing users per BS. The method is a ranking-based per-antenna selection that preserves multiuser diversity in CBF systems and reduces the amount of information exchanged between the BSs and the CU. Numerical results show that our proposed metrics and algorithms for user selection can achieve a large portion of the optimal sum rate (the benchmark is a fully centralized system) by exploiting local CSI with limited message exchange between BSs and the CU.

The remainder of the paper is organized as follows. The system model and the problem formulation are presented in Section II. In Section III we present the DZF and DVSINR precoding schemes, their properties, expressions for their effective channel gains, and their relation with user selection. In Section IV we define the metrics that estimate the effective channel gains and Section V presents the semi-distributed user selection algorithm whose solution set solves the sum rate maximization problem. Numerical results are provided in Section VI and conclusions are drawn in Section VII.

Notation: matrices and vectors are set in upper and lower boldface respectively. $\langle \mathbf{a}, \mathbf{b} \rangle$ is the inner product between vectors \mathbf{a} and \mathbf{b} . $(\cdot)^T$, $(\cdot)^H$, $|\cdot|$, $\|\cdot\|$ denote the transpose, hermitian transpose, absolute value, and vector norm respectively. Calligraphic letters, e.g. \mathcal{G} , denote sets and $|\mathcal{G}|$ denotes cardinality. $Tr(\cdot)$, and $\det(\cdot)$ represent the trace and determinant operators. $\mathbb{E}[\cdot]$ represents the expectation operation. $Sp(\mathbf{A})$ and $Sp(\mathbf{A})^\perp$ denote the subspace and orthogonal subspace spanned by the columns of matrix \mathbf{A} . $\lambda_i(\mathbf{A})$ is the i th eigenvalue of the operated matrix, $\lambda_{\max}(\mathbf{A})$, $\lambda_{\min}(\mathbf{A})$, $rank(\mathbf{A})$, $null(\mathbf{A})$ are the maximum and minimum eigenvalues, rank and null space of matrix \mathbf{A} respectively. $\mathbf{eig}(\mathbf{A}) = [\lambda_1, \dots, \lambda_n]$ is the vector that contains all n

eigenvalues of matrix \mathbf{A} . Let \mathbf{x} be a vector, then $[\mathbf{x}]_i = x_i$ is the i th element. \mathbf{I}_n is the identity matrix of size n . \mathbb{R}_+ is the set of nonnegative real numbers. For a given a vector $\mathbf{x} \in \mathbb{R}_+^n$, the Jain's index of fairness is defined as follows [23]:

$$J(\mathbf{x}) \triangleq \frac{(\sum_{i=1}^n x_i)^2}{n \sum_{i=1}^n x_i^2}, \quad (1)$$

where $\{J(\cdot) \in \mathbb{R}_+ | \frac{1}{n} \leq J(\cdot) \leq 1\}$.

II. PROBLEM FORMULATION

We consider a multiuser multicell clustered network where a group of B adjacent BSs form a cluster. Each BS has N_t antennas, all users in the network are equipped with single antenna devices, and define $\epsilon \triangleq \max\{N_t - (B - 1), 0\}$. The BSs only exchange messages of scheduling control through a CU and precoding design is performed at each BS using local CSI. The joint user selection and precoding design are performed for cell-edge users located in the cell-edge area defined by B BSs. The users are deployed within a circular area that spans a radius r_{coop} (a fraction of the cell radius r). The b th BS has one index set of edge users \mathcal{S}_b and it only transmits data to one user in this set. Consider that $\mathcal{S}_b \cap \mathcal{S}_j = \emptyset, \forall j \neq b$ and the transmitted signal from BS b to user $k_b \in \mathcal{S}_b$ is: $\mathbf{x}_b = \sqrt{P_b} \mathbf{w}_b s_b$. P_b is the transmitted power, $\mathbf{w}_b \in \mathbb{C}^{N_t \times 1}$ is the unit norm precoder and s_b is the transmitted data symbol with $\mathbb{E}[|s_b|^2] = 1$, $\mathbb{E}[\|\mathbf{x}_b\|^2] = P_b$, and $P_b \leq P$ where P is the maximum available power. The received signal of the intended user k_b of BS b is given by:

$$y_{bk_b} = \sqrt{P_b} \mathbf{h}_{bk_b}^H \mathbf{w}_b s_b + \sum_{j=1, j \neq b}^B \sqrt{P_j} \mathbf{h}_{jk_b}^H \mathbf{w}_j s_j + n_{k_b}, \quad (2)$$

where $\mathbf{h}_{bk_b} \sim \mathcal{CN}(0, \varrho_{bk_b}^2 \mathbf{I})$ of size $N_t \times 1$ is a flat Rayleigh fading propagation channel between user k_b and BS b and $\varrho_{bk_b}^2$ is the long-term channel power gain. The term $n_{k_b} \sim \mathcal{CN}(0, \sigma_n^2)$ is the noise. The receivers treat co-terminal interference as noise and the instantaneous signal-to-interference-plus-noise ratio (SINR) of user $k_b \in \mathcal{S}_b$ is defined as:

$$SINR_{bk_b} = \frac{P_b |\mathbf{h}_{bk_b}^H \mathbf{w}_b|^2}{\sum_{j=1, j \neq b}^B P_j |\mathbf{h}_{jk_b}^H \mathbf{w}_j|^2 + \sigma_n^2}. \quad (3)$$

In a cluster with B BSs, there exists $L = \prod_{b=1}^B |\mathcal{S}_b|$ user permutations of B users that can be chosen for simultaneous transmission. Each user in $\mathcal{S} = \bigcup_{b=1}^B \mathcal{S}_b$ has a unique index and all BSs

know which indices belong to each BS. Let $\mathcal{G}_l \forall l \in \{1, \dots, L\}$ be a set of B users where each user is served by one BS and the users indices in the set l are the same for all BSs. The set \mathcal{G}_l has an associated channel matrix at the b th BS which is given by $\mathbf{H}_b^{(l)} \triangleq \{\mathbf{h}_{bk_i} : k_i \in \mathcal{G}_l\}$, i.e., all the local channels of the users grouped in \mathcal{G}_l . We need to solve the sum rate maximization problem in the multiuser multicell scenario defined as:

$$\begin{aligned} & \underset{l \in \{1, \dots, L\}}{\text{maximize}} \quad \sum_{b=1}^B \log_2 (1 + SINR_{bk_b}) \\ & \text{subject to} \quad \|\mathbf{w}_b^{(type)}(\mathbf{H}_b^{(l)})\|^2 = 1, \quad \forall b \in \{1, \dots, B\} \end{aligned} \quad (4)$$

where $SINR_{bk_b}$ is defined in (3) for the user $k_b \in \{\mathcal{S}_b \cap \mathcal{G}_l\}$. The precoding vectors $\mathbf{w}_b^{(type)}(\mathbf{H}_b^{(l)}) \forall b$ are functions of $\mathbf{H}_b^{(l)}$ at each BS for the given set l , and $type \in \{DZF, DVSINR\}$ is the implemented precoding technique which will be defined in the next section. Our objective is to find the set l that solves problem (4) which can be attained by taking advantage of the properties of $\mathbf{w}_b^{(type)}$. Such properties are used to exploit the local CSI in order to evaluate the effective channel gains, i.e., $|\mathbf{h}_{bk_b}^H \mathbf{w}_b|^2$ which are tightly related with the achievable rates.

III. DISTRIBUTED LINEAR PRECODING

In this section we investigate two precoding techniques DZF and DVSINR. We need to define underlying characteristics of the precoders, their dependence on the SNR regime and N_t , and quantify how those characteristics affect the instantaneous and average effective channel gains.

A. Distributed Zero Forcing (DZF)

Zero-forcing is a precoding strategy that removes the inter-user interference and is defined always that $N_t \geq B$. The conditions to achieve near Pareto-optimal rates with distributed ZF for the two-BS scenarios were presented in [14] and for B BSs generalized expressions to compute \mathbf{w}_b are provided in [6], [12]. Let $\tilde{\mathbf{H}}_{bk_b}$ be the aggregate interference matrix of user k_b given by:

$$\tilde{\mathbf{H}}_{bk_b} = [\mathbf{h}_{bk_1}, \dots, \mathbf{h}_{bk_{b-1}}, \mathbf{h}_{bk_{b+1}}, \dots, \mathbf{h}_{bk_B}], \quad (5)$$

and each term $\mathbf{h}_{bk_i} \forall i \neq b$ corresponds to the channel between BS b and the non-intended user $k_i \in \mathcal{S}_i$. The matrix $\tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk_b}} = \text{null}(\tilde{\mathbf{H}}_{bk_b})$ contains ϵ column vectors¹ that are candidates to form \mathbf{w}_b since they will produce zero interference to the other users in $\tilde{\mathbf{H}}_{bk_b}$. If $\epsilon > 1$ the elements of $\tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk_b}}$ can be linearly combined to form the precoding vector as follows [12]:

$$\mathbf{w}_b^{(DZF)} = \tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk_b}} \frac{(\mathbf{h}_{bk_b}^H \tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk_b}})^H}{\|\mathbf{h}_{bk_b}^H \tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk_b}}\|}, \quad (6)$$

and the received signal at user k_b has its phase aligned.

Proposition 1. *The expected value of the effective channel gain of the intended user $k_b \in \mathcal{S}_b$ served by BS b using DZF precoding with \mathbf{w}_b defined in (6) under constraint $N_t \geq B$ is defined as follows:*

$$\mathbb{E} [|\mathbf{h}_{bk_b}^H \mathbf{w}_b|^2] = \frac{\epsilon}{N_t} \mathbb{E} [\|\mathbf{h}_{bk_b}\|^2] \quad (7)$$

Proof: See Appendix A ■

B. Distributed Virtual SINR (DVSINR)

The ideal precoder technique would be able to balance between signal power maximization and interference power minimization and a heuristic way to find such balance is reached by maximizing the SLNR [6]. In [8] the authors show that it is possible to achieve Pareto-optimal rates in multicell transmission when the precoding vectors are given by:

$$\mathbf{w}_b^* = \arg \max_{\|\mathbf{w}\|^2=1} \frac{v_{bk_b} |\mathbf{h}_{bk_b}^H \mathbf{w}|^2}{\sum_{j=1, j \neq b}^B v_{bk_j} |\mathbf{h}_{bk_j}^H \mathbf{w}|^2 + \frac{\sigma_n^2}{P_b}}, \quad (8)$$

where $v_{bk_b} \in (0, 1)$ which is a heuristic extension of the SLNR precoding [6], [22]. Then $\mathbf{w}_b^* \forall b$ are linear combinations of the maximal ratio transmission and ZF precoders and the coefficients v_{bk_b} that optimally maximize the sum rate can be only computed with global CSI. If maximum ICI is accounted² $v_{bk_b} = 1 \forall b$ the precoders that solve the virtual SINR maximization problem (8) are given by [8]:

¹Using SVD $\tilde{\mathbf{H}}_{bk_b} = [\tilde{\mathbf{W}}_{\tilde{\mathbf{H}}_{bk_b}} \tilde{\mathbf{W}}_{\tilde{\mathbf{H}}_{bk_b}}] \Sigma_{\tilde{\mathbf{H}}_{bk_b}} \mathbf{O}_{\tilde{\mathbf{H}}_{bk_b}}^H$ where $\tilde{\mathbf{W}}_{\tilde{\mathbf{H}}_{bk_b}}$ contains ϵ orthonormal vectors that are the basis of the null space of $\tilde{\mathbf{H}}_{bk_b}$ and $\tilde{\mathbf{V}}_{\tilde{\mathbf{H}}_{bk_b}} = \text{null}(\tilde{\mathbf{H}}_{bk_b}) = \tilde{\mathbf{W}}_{\tilde{\mathbf{H}}_{bk_b}}$.

²The authors in [5] showed that the coefficients v_{bk} can define user weights that may represent, for instance, user priority.

$$\mathbf{w}_b^{(DVSINR)} = \frac{\mathbf{D}_{bk_b} \mathbf{h}_{bk_b}}{\|\mathbf{D}_{bk_b} \mathbf{h}_{bk_b}\|}, \quad (9)$$

where $\mathbf{D}_{bk_b} = \mathbf{C}_{bk_b}^{-1}$, $\mathbf{C}_{bk_b} \triangleq \rho_b^{-1} \mathbf{I}_{N_t} + \tilde{\mathbf{H}}_{bk_b} \tilde{\mathbf{H}}_{bk_b}^H$ is a $N_t \times N_t$ positive-definite Hermitian matrix and $\rho_b = \frac{P_b}{\sigma_n^2}$. The following result describes the relation between the eigenvalues of \mathbf{D}_{bk_b} and the expected value of the effective channel gain.

Proposition 2. *The effective channel gain of the user $k_b \in \mathcal{S}_b$ served by the b th BS under DVSINR precoding constrained by $N_t \geq B$ can be bounded as follows:*

$$\mathbb{E} [|\mathbf{h}_{bk_b}^H \mathbf{w}_b|^2] \approx \mathbb{E} [\|\mathbf{h}_{bk_b}\|^2 J(\mathbf{eig}(\mathbf{D}_{bk_b}))] \quad (10)$$

where \mathbf{w}_b is defined in (9) and $J(\cdot)$ is the Jain's fairness index.

Proof: See Appendix B ■

Proposition 3. *For DVSINR, given the matrix $\tilde{\mathbf{H}}_{bk_b} \in \mathbb{C}^{N_t \times (B-1)}$ and its corresponding $\mathbf{D}_{bk_b} \in \mathbb{C}^{N_t \times N_t}$ under constraint $N_t \geq B$ it holds that*

$$\lim_{\rho_b \rightarrow \infty} \left(J(\mathbf{eig}(\mathbf{D}_{bk_b})) - \frac{\epsilon}{N_t} \right) = 0, \quad (11)$$

which implies that $\exists \rho_0$ and $\forall \rho_b \geq \rho_0$ the expected value of the effective channel gain is upper bounded as follows:

$$\mathbb{E} [|\mathbf{h}_{bk_b}^H \mathbf{w}_b|^2] \leq \frac{\epsilon}{N_t} \mathbb{E} [\|\mathbf{h}_{bk_b}\|^2] \quad (12)$$

Proof: See Appendix C ■

The ICI for DVSINR is nonzero and for the high SNR regime the interference components in the denominator of (3) are usually neglected [5], [8]. The following result provides an approximation of the power that is leaked from clustered BSs using DVSINR precoding.

Proposition 4. *For DVSINR and $N_t \geq B$, the magnitude of the interference or leakage from the j th BS over the channel $\mathbf{h}_{jk_b} \in \tilde{\mathbf{H}}_{jk_j} \forall j \neq b$ in the denominator of (3) for the user $k_b \in \mathcal{S}_b$ served by the b th BS can be approximated as follows:*

$$\mathbb{E} [|\mathbf{h}_{jk_b}^H \mathbf{w}_j|^2] \approx \mathbb{E} \left[\frac{\|\mathbf{h}_{jk_b}\|^2}{\epsilon(\rho_j \lambda_{\min}(\tilde{\mathbf{H}}_{jk_j}^H \tilde{\mathbf{H}}_{jk_j}) + 1)^2} \right], \quad (13)$$

where \mathbf{w}_j is a function of the matrix $\tilde{\mathbf{H}}_{jk_j}$ associated to the user k_j served by the j th BS.

Proof: See Appendix C ■

C. Distributed Linear Precoding and User Selection

Consider that $k_b \in \{\mathcal{S}_b \cap \mathcal{G}_l\}$ and let $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l) \in \mathbb{C}^{N_t \times (B-1)}$ be the aggregate interference matrix of k_b which contains all channels of $\mathbf{H}_b^{(l)}$ except \mathbf{h}_{bk_b} .

1) *DZF*: This scheme is defined if $N_t \geq B$ and achieves zero inter-user interference, i.e., $|\mathbf{h}_{bk_j}^H \mathbf{w}_b^{(DZF)}|^2 = 0, \forall k_j \in \mathcal{G}_l \setminus \{k_b\}$. From Proposition 1 observe that the average value of $|\mathbf{h}_{bk_b}^H \mathbf{w}_b^{(DZF)}|^2$ depends on N_t and ϵ . As ϵ grows the effective channel gain is enhanced. However, results in Appendix A show that the instantaneous effective channel gain is a function of the angle between \mathbf{h}_{bk_b} and the basis of the null subspace of $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$. The system performance is optimized, regardless the SNR regime, if the intended direct channel \mathbf{h}_{bk_b} at BS b maximize $|\mathbf{h}_{bk_b}^H \mathbf{w}_b^{(DZF)}|^2$. This means that channel magnitude and spatial compatibility (quasi-orthogonality w.r.t. $Sp(\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l))$) must be optimized jointly. Notice that $\forall k_b \in \mathcal{S}_b$ there exists $\prod_{j=1, j \neq b}^B |\mathcal{S}_j|$ possible precoders and the set \mathcal{G}_l that maximizes $|\mathbf{h}_{bk_b}^H \mathbf{w}_b^{(DZF)}|^2$ at BS b is, in general, not the best set at other BSs.

2) *DVSINR*: This scheme does not impose a constraint on N_t but its capacity to combat inter-user interference depends on it. For a given user set \mathcal{G}_l , Proposition 2 shows that in the low and medium SNR regimes the expected value of $|\mathbf{h}_{bk_b}^H \mathbf{w}_b^{(DVSINR)}|^2$ depends on the magnitude of \mathbf{h}_{bk_b} and the characteristics of $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$. In particular, the magnitude of each i.i.d. vector in $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$ and its singular values which directly modifies $J(\text{eig}(\mathbf{D}_{bk_b}))$. In the high SNR, Proposition 3 indicates that the expected value of $|\mathbf{h}_{bk_b}^H \mathbf{w}_b^{(DVSINR)}|^2$ is limited by ϵ and N_t similar to DZF. Since the impact of \mathbf{D}_{bk_b} in the effective channel gain is dominated by ϵ eigenvalues associated to the orthogonal subspace to $\mathcal{V}_{bk_b} = Sp(\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l))$, the selected user at each BS should meet the same conditions previously described for DZF. At the low SNR the eigenvalues of \mathbf{D}_{bk_b} have similar magnitudes³ and the BS can select its user selfishly based on the channel magnitudes regardless the characteristics of \mathcal{V}_{bk_b} . At medium SNR the user selection is more complicated since the instantaneous effective channel gain is modified by the weighted basis of \mathcal{V}_{bk_b} where the weights are functions of ρ_b and ϵ , cf. (32) in Appendix B. Notice that because $|\mathbf{h}_{bk_j}^H \mathbf{w}_b^{(DVSINR)}|^2 \neq 0, \forall k_j \in \mathcal{G}_l \setminus \{k_b\}$ the achievable SINR (3) strongly depends on N_t and B . If $\epsilon > 0$ (power limited

³From the definition in (34) observe that for the low SNR regime $\rho^{-1} > [\Sigma_{\hat{\mathbf{H}}}]_{ii}$, i.e., the eigenvalues of $\hat{\mathbf{H}}$ are negligible.

scenario) the amount of leaked power from BS j to the user k_b served by BS b is scaled by a factor ϵ^{-1} as shown in (13). When $\rho_j \rightarrow \infty$ the leakage is also scaled by a factor of ρ_j^{-2} according to Proposition 4 and inter-user interference vanishes. The expression (13) reveals that for a fixed ρ_j the leakage is minimized if $\lambda_{\min}(\tilde{\mathbf{H}}_{jk_j}^H \tilde{\mathbf{H}}_{jk_j})$ is maximized, which occurs if the i.i.d. vectors in $\tilde{\mathbf{H}}_{jk_j}$ are quasi-orthogonal. For user selection purposes, at BS b the best set \mathcal{G}_l should meet two conditions: 1) \mathbf{h}_{bk_b} is quasi-orthogonal to \mathcal{V}_{bk_b} (similar to DZF), and 2) the elements in $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$ are quasi-orthogonal. If $\epsilon = 0$ (interference limited scenario) a strategy for user selection based only on local CSI is hard to define because the channels of all user in \mathcal{G}_l are coupled in the SINR expression (3). In other words, accurate user selection in such scenario requires CSI exchange between BSs.

IV. METRICS OF SPATIAL COMPATIBILITY

In this section we answer the question: *what kind of information can be extracted from the local CSI and sent to the CU in order to perform scheduling?*. We define channel metrics whose objective is to measure spatial compatibility between users taking into account the SNR regime and N_t .

A. Power Limited Scenario: $N_t \geq B$

Due to the fact that global CSI is not available at the CU, centralized user selection (e.g., [2], [11], [24]) cannot be performed. In order to design semi-distributed user selection we need to define the type of scheduling control information exchanged between the BSs and the CU. We say that metric g_{bl} is a function of the local CSI $\mathbf{H}_b^{(l)}$ so that $g_{bl} : \mathbb{C}^{N_t \times B} \mapsto \mathbb{R}_+$. Such mapping computes an approximation of $|\mathbf{h}_{bk_b}^H \mathbf{w}_b|^2$, i.e., it quantifies how profitable is to select the set \mathcal{G}_l for transmission at the b th BS. Let $\mathbf{P}_{\mathbf{h}_{bk_b}} = \tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)(\tilde{\mathbf{H}}_{bk_b}^H(\mathcal{G}_l)\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l))^{-1}\tilde{\mathbf{H}}_{bk_b}^H(\mathcal{G}_l)$ be the projector matrix onto \mathcal{V}_{bk_b} , and $\mathbf{Q}_{\mathbf{h}_{bk_b}} = \mathbf{I}_{N_t} - \mathbf{P}_{\mathbf{h}_{bk_b}}$ the projector matrix onto the orthogonal complement of \mathcal{V}_{bk_b} [25]. The proposed metric to estimate $|\mathbf{h}_{bk_b}^H \mathbf{w}_b|^2$ is given by:

$$g_{bl} = \|\mathbf{Q}_{\mathbf{h}_{bk_b}} \mathbf{h}_{bk_b}\|^2 + \alpha_{bk_b} \|\mathbf{P}_{\mathbf{h}_{bk_b}} \mathbf{h}_{bk_b}\|^2, \quad (14)$$

where α_{bk_b} is a function of the type of precoding scheme.

For DZF $\alpha_{bk_b} = 0$ for all ρ_b since the precoder takes the form $\mathbf{w}_b^{(DZF)} = \mathbf{Q}_{\mathbf{h}_{bk_b}} \mathbf{h}_{bk_b} / \|\mathbf{Q}_{\mathbf{h}_{bk_b}} \mathbf{h}_{bk_b}\|$ which is the direction of the projection of \mathbf{h}_{bk_b} onto $Sp(\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l))^\perp$. For the case of DVSINR,

$\mathbf{w}_b^{(DVSINR)} = \mathbf{h}_{bk_b} / \|\mathbf{h}_{bk_b}\|$ as $\rho_b \rightarrow 0$, i.e., the precoder is given by the matched filter and we must have $\alpha_{bk_b} = 1$ in order to meet $g_{bl} = |\mathbf{h}_{bk_b}^H \mathbf{w}_b^{(DVSINR)}|^2 = \|\mathbf{h}_{bk_b}\|^2$. When $\rho_b \rightarrow \infty$ the precoder is given by $\mathbf{w}_b^{(DVSINR)} = \mathbf{Q}_{\mathbf{h}_{bk_b}} \mathbf{h}_{bk_b} / \|\mathbf{Q}_{\mathbf{h}_{bk_b}} \mathbf{h}_{bk_b}\|$ and we must have $\alpha_{bk_b} = 0$ so that $g_{bl} = |\mathbf{h}_{bk_b}^H \mathbf{w}_b^{(DVSINR)}|^2 = \|\mathbf{Q}_{\mathbf{h}_{bk_b}} \mathbf{h}_{bk_b}\|^2$. Therefore, α_{bk_b} must change its value depending on the SNR regime and the characteristics of the i.i.d. vectors in $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$.

Proposition 5. *The instantaneous effective channel gain $|\mathbf{h}_{bk_b}^H \mathbf{w}_b^{(DVSINR)}|^2$ is given by a nonlinear combination of the orthonormal basis of both $Sp(\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l))$ and $Sp(\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l))^\perp$. The metric (14) is an approximation of $|\mathbf{h}_{bk_b}^H \mathbf{w}_b^{(DVSINR)}|^2$ and a heuristic definition of the weight α_{bk_b} is given by:*

$$\alpha_{bk_b} = \frac{1}{(\rho_b \lambda_{\max}(\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)^H \tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)) + 1)^2}. \quad (15)$$

Proof: See Appendix D ■

B. Interference Limited Scenario: $N_t < B$

In this scenario $\epsilon = 0$, DZF is not defined [8], and DVSINR precoding can be implemented but inter-user interference is unavoidable. Moreover, metric (14) does not provide information for user selection or cannot be computed. If $B - N_t = 1$ then $\mathbf{P}_{\mathbf{h}_{bk_b}} = \mathbf{I}_{N_t}$ and we cannot extract useful information from (14). If $B - N_t > 1$ the matrix $\tilde{\mathbf{H}}_{bk_b}^H(\mathcal{G}_l) \tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$ is ill-conditioned⁴ and $\mathbf{P}_{\mathbf{h}_{bk_b}}$ is no longer a projector matrix. Therefore, we want to define a metric of the form $g_{bl} = \|\mathbf{h}_{bk_b}\|^2 f(\mathbf{H}_b^{(l)}, \rho_b)$. The function $f(\mathbf{H}_b^{(l)}, \rho_b)$ must become 1 as $\rho_b \rightarrow 0$ whilst its value should change according to the strength of \mathbf{h}_{bk_b} and its spatial relation with \mathcal{V}_{bk_b} as $\rho_b \rightarrow \infty$.

Consider two grouped users $k_b \in \{\mathcal{S}_b \cap \mathcal{G}_l\}$, $k_j \in \{\mathcal{S}_j \cap \mathcal{G}_l\}$, and define $m_b^{(k_b k_j)} = \|\mathbf{h}_{bk_b}\|^2 / \|\mathbf{h}_{bk_j}\|^2$. $m_b^{(k_b k_j)}$ can provide a coarse estimation of the location of the users regarding the b th BS. If $m_b^{(k_b k_j)} \approx 1$ this may suggest either that k_b and k_j are close to each other at the cell-edge, or that k_b is far from BS b and transmission over channel \mathbf{h}_{bk_b} could be affected by strong interference. For $m_b^{(k_b k_j)} \gg 1$, k_b may be close to BS b or k_j is either far from BS b or experiencing deep fading. If $m_b^{(k_b k_j)} < 1$ fading is large in \mathbf{h}_{bk_b} and transmission may be infeasible. In order to quantify how strong and reliable for transmission is \mathbf{h}_{bk_b} using local CSI, define the coefficient $M_{bk_b}^{(l)}$ as:

⁴Observe that $\tilde{\mathbf{H}}_{bk_b}^H(\mathcal{G}_l) \tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$ is a matrix of size $B - 1 \times B - 1$ which has N_t non-zero eigenvalues. When $B - N_t > 1$ the ratio $\lambda_{\max}(\tilde{\mathbf{H}}_{bk_b}^H(\mathcal{G}_l) \tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)) / \lambda_{\min}(\tilde{\mathbf{H}}_{bk_b}^H(\mathcal{G}_l) \tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)) \rightarrow \infty$ and the matrix is close to singular [26].

$$M_{bk_b}^{(l)} = \frac{\|\mathbf{h}_{bk_b}\|^2}{\prod_{j=1, j \neq b}^B \|\mathbf{h}_{bk_j}\|^{2/(B-1)}}, \quad (16)$$

where the denominator is the geometric mean of the squared norms of the column vectors of $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$. Using the geometric mean has two objectives: collecting in a single quantity the strength of the channels $\{\mathbf{h}_{bk_j}\}_{k_j \in \mathcal{G}_l \setminus \{k_b\}}$ and considering the effects of each magnitude equally⁵ in the averaging operation.

We also need to estimate the spatial compatibility between all the elements of $\mathbf{H}_b^{(l)}$, the degradation due to correlation in $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$ and the effects of ρ_b . In other words, we need an operation similar to the NSP. Define the metric for spatial compatibility as

$$\zeta_{bk_b}^{(l)} = \frac{\left| \det \left((\mathbf{H}_b^{(l)})^H \mathbf{H}_b^{(l)} \right) \right|}{\left| \det \left(\rho_b^{-1} \mathbf{I}_{N_t} + \tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l) (\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l))^H \right) \right|}, \quad (17)$$

which is the ratio between the volume of a $B \times B$ matrix over the the volume of a $N_t \times N_t$ matrix. Recall that the determinant measures the volume spanned by the columns of a matrix. The more orthogonal the column vectors of a matrix, the larger the value of its determinant [26].

The heuristic metric for user selection is defined as

$$g_{bl} = \|\mathbf{h}_{bk_b}\|^2 \left(\alpha_{bk_b} + (1 - \alpha_{bk_b}) M_{bk_b}^{(l)} \zeta_{bk_b}^{(l)} \right), \quad (18)$$

where α_{bk_b} is given by (15). Observe that in the low SNR $\alpha_{bk_b} \rightarrow 1$ which yields $g_{bl} \approx \|\mathbf{h}_{bk_b}\|^2$. In the high SNR $\alpha_{bk_b} \rightarrow 0$ and the selection metric is $g_{bl} \approx \|\mathbf{h}_{bk_b}\|^2 M_{bk_b}^{(l)} \zeta_{bk_b}^{(l)}$.

C. NSP Approximation

The NSP operation $\|\mathbf{Q}_{\mathbf{h}_{bk_b}} \mathbf{h}_{bk_b}\|^2 = \|\mathbf{h}_{bk_b}\|^2 \sin^2 \theta_{\mathcal{V}_{bk_b} \mathbf{h}_{bk_b}}$ can be approximated using the inner products of the elements of $\mathbf{H}_b^{(l)}$ which reduces the number of arithmetic operations required to compute metrics (14) or (18). The term $\theta_{\mathcal{V}_{bk_b} \mathbf{h}_{bk_b}}$ is the angle between \mathbf{h}_{bk_b} and the subspace \mathcal{V}_{bk_b} . For two i.i.d. channels \mathbf{h}_{bk_b} and \mathbf{h}_{bi_b} at the b th BS, the spatial compatibility between them can be measured by the coefficient of correlation defined as [25]:

$$\eta_{\mathbf{h}_{bk_b} \mathbf{h}_{bi_b}} = \frac{|\langle \mathbf{h}_{bk_b}, \mathbf{h}_{bi_b} \rangle|}{\|\mathbf{h}_{bk_b}\| \|\mathbf{h}_{bi_b}\|}, \quad (19)$$

⁵This is not the case for the arithmetic mean since the magnitudes $\{\|\mathbf{h}_{bk_j}\|^2\}_{k_j \in \mathcal{G}_l \setminus \{k_b\}}$ may have a large variance in which case the smallest magnitudes would be neglected.

where the coefficient $0 \leq \eta_{\mathbf{h}_{bk_b} \mathbf{h}_{bi_b}} = \cos \theta_{\mathbf{h}_{bk_b} \mathbf{h}_{bi_b}} \leq 1$ geometrically represents the cosine of the angle between the two channel vectors. The coefficient $\sin^2 \theta_{\mathbf{h}_{bk_b} \mathbf{h}_{bi_b}}$ that scales $\|\mathbf{h}_{bk_b}\|^2$ in a NSP operation can be computed as [25]:

$$\sin^2 \theta_{\mathbf{h}_{bk_b} \mathbf{h}_{bi_b}} = (1 - \eta_{\mathbf{h}_{bk_b} \pi(1)}^2) \dots (1 - \eta_{\mathbf{h}_{bk_b} \pi(i)|\pi(1)\dots\pi(i-1)}^2), \quad (20)$$

where $\pi(i)$ is the i th ordered element of $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$ and $\eta_{\mathbf{h}_{bk_b} \pi(i)|\pi(1)\dots\pi(i-1)}$ is the partial correlation coefficient between the channel vector \mathbf{h}_{bk_b} and the selected vector associated with $\pi(i)$ eliminating the effects due to the previous ordered elements $\pi(1), \pi(2), \dots, \pi(i-1)$.

If the correlation coefficients (19) are used instead of the partial correlation coefficients in (20) a suboptimal evaluation of $\sin^2 \theta_{\mathbf{h}_{bk_b} \mathbf{h}_{bi_b}}$ can be computed. Using this approximation of the NSP, the reported metric to the CU by the b th BS for the user $k_b \in \{\mathcal{S}_b \cap \mathcal{G}_l\}$ is given by:

$$g_{bl} = \|\mathbf{h}_{bk_b}\|^2 \prod_{i_b \neq k_b, i_b \in \mathcal{G}_l} \sin^2 \theta_{\mathbf{h}_{bk_b} \mathbf{h}_{bi_b}}. \quad (21)$$

Observe that metric (21) can be computed even if $N_t < B$ since (19) is independent of B and exists for all $N_t \geq 2$. If $N_t \geq B$ metric (21) is an upper bound of the NSP. This means that $\|\mathbf{h}_{bk_b}\|^2$ is scaled by a coefficient larger than $\sin^2 \theta_{\mathbf{h}_{bk_b} \mathbf{h}_{bi_b}}$ which prioritizes the channel magnitude over the spatial compatibility when the user selection is performed. The relationship between the real and the approximated expected values of the NSP is presented in the following proposition.

Proposition 6. *For $N_t \geq B$ it holds that the average value of metric (21) is an upper bound of the average metric (14) with $\alpha_{bk_b} = 0$, i.e., the NSP, so that:*

$$\mathbb{E} \left[\|\mathbf{h}_{bk_b} \mathbf{Q}_{\mathbf{h}_{bk_b}}\|^2 \right] \leq \mathbb{E} \left[\|\mathbf{h}_{bk_b}\|^2 \prod_{i_b \neq k_b, i_b \in \mathcal{G}_l} \sin^2 \theta_{\mathbf{h}_{bk_b} \mathbf{h}_{bi_b}} \right] \quad (22)$$

Proof: See Appendix D ■

V. THE MULTICELL USER SELECTION

Once that g_{bl} has extracted spatial compatibility information from the multiuser channel matrix $\mathbf{H}_b^{(l)}$ we need to answer two questions: 1) *what should be the optimization over the metrics g_{bl} at the CU in order to find the set \mathcal{G}_{l^*} of most spatially compatible users?* and 2) *how to minimize the number of metrics g_{bl} computed per BS so that \mathcal{G}_{l^*} achieves close-to-optimal performance and multiuser diversity is preserved?*

A. Exhaustive Search Selection over the Metrics

The optimal solution of (4) can be only found by exhaustive searching over the achievable rates of the sets $\mathcal{G}_l \forall l \in \{1, \dots, L\}$. Such a search requires global CSI at the CU and the computation of BL precoders in order to accurately evaluate the L possible achievable sum rates. A sub-optimal solution to (4) can be found by avoiding the full CSI exchange with the CU and instead reporting the metrics computed by (14), (18), or (21). Assuming that all BSs know the L ordered sets, the b th BS computes the metrics $g_{bl} \forall l$ and report them to the CU where the index of the set that is chosen to perform coordinated transmission is found solving the following problem:

$$l^* = \arg \max_{l \in \{1, \dots, L\}} \prod_{b=1}^B g_{bl}. \quad (23)$$

Bearing in mind that g_{bl} attempts to estimate the effective channel gains, the rationale behind the product in (23) is that for MISO transmission a set of users maximizing the product of their effective channel gains also achieves maximum sum rate [18]. In our scenario, taking the product of the metrics assigns the same priority to each independent metric $g_{bl} \forall b$. This means that the computation of l^* is not biased by a dominant metric $g_{bl} \gg g_{jl} \forall j \neq b$ for a given set l , which would be only beneficial to BS b . Once that l^* has been found, the BSs use the matrices $\mathbf{H}_b^{(l^*)} \forall b$ to locally compute the precoders which are used to sub-optimally solve (4).

B. Search Space Pruning

Previously we have discussed that solving (23) does not require global CSI but L metrics are reported from each BS to the CU. If the number of cell-edge users is large ($|\mathcal{S}| \gg BN_t$) computing the metrics for all user permutations L may become prohibitive. In single-cell systems the authors in [27] showed that for fixed N_t and single-antenna users, the system capacity under spatial division multiple access scales by $N_t \log(\log(|\mathcal{S}_b|))$ at the b th BS. This result means that multiuser diversity provides a marginal contribution to the capacity enhancement unless $|\mathcal{S}_b| \rightarrow \infty$. Similar conclusions extend to multi-cell systems operating in JT mode (e.g., [9], [24]). Numerical results in [9] show that multiuser diversity is beneficial for BS cooperation when only a fraction of the total number of users is considered to participate in the selection process. For a multi-cell JT system employing ZF precoding [24], BN_t transmit antennas can serve at most

the same number of single-antenna users, and low-complexity user selection algorithms can be extended from single-cell systems [16]–[20].

In our CBF scenario we want: to achieve multiplexing gain; decrease the solution space's size of problem (23) by selecting a small fraction of competing users from \mathcal{S} ; and preserve multiuser diversity when selecting the competing users. In order to find a subset $\hat{\mathcal{S}}_b \subseteq \mathcal{S}_b$ at BS b , let $\mathbf{h}_{bk_b} = [h_{bk_b1}, \dots, h_{bk_bN_t}]^T$ be the channel of the user $k_b \in \mathcal{S}_b$ where h_{bk_bn} is the channel component of the n th antenna. Consider the following: 1) For DZF efficient user selection must be focused on finding quasi-orthogonal users regarding the SNR regime. 2) For DVSINR efficient user selection in the low SNR is determined by the channel magnitude. 3) In the high SNR the effective channel gains of DZF and DVSINR are similar and efficient user selection must find spatially orthogonal users. A fast way to find a set of quasi-orthogonal users in JT systems is by applying a ranking-based per-antenna selection as in [24]. The idea behind such selection is that for two user, k_b and i_b having $|h_{bk_bn}| > |h_{bk_bn'}| \forall n' \neq n$, $|h_{bi_bm}| > |h_{bi_bm'}| \forall m' \neq m$, and $\forall n \neq m$, the inner product $\langle \mathbf{h}_{bk_b}, \mathbf{h}_{bi_b} \rangle$ decreases as the magnitude of each dominant antenna n and m increases, i.e., they become quasi-orthogonal.

In our scenario we require the channel of the selected user $k_b \in \{\mathcal{S}_b \cap \mathcal{G}_l\}$ of BS b to be as orthogonal as possible w.r.t. the channels in $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$. Therefore, the per-antenna ranking can be used for pre-selecting the users with maximum per-antenna channel magnitude. In this way a user $k_b \in \hat{\mathcal{S}}_b$ will have a dominant antenna (spatial direction) n and it is likely that channels in $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$ do not have per-antenna channel magnitudes similar or closed to $|h_{bk_bn}|$ at the same antenna n due to path-loss effects, which provides certain degree of spatial compatibility.

Define the dominant user for the antenna n at BS b as

$$k_{b(n)} = \arg \max_{i_b \in \mathcal{S}_b} |h_{bi_bn}|, \quad (24)$$

and let the user with the largest channel magnitude be

$$k_{b(\max)} = \arg \max_{i_b \in \mathcal{S}_b} \|\mathbf{h}_{bi_b}\|, \quad (25)$$

where the subset of users that will participate in the selection process at the b th BS is defined as

$$\hat{\mathcal{S}}_b = \{k_{b(n)}\}_{n=1}^{N_t} \cup \{k_{b(\max)}\}. \quad (26)$$

This user pre-selection reduces the size of the search space because it only considers the strongest users per spatial direction per BS. Including $k_{b(\max)}$ in the set $\hat{\mathcal{S}}_b$ guarantees that for

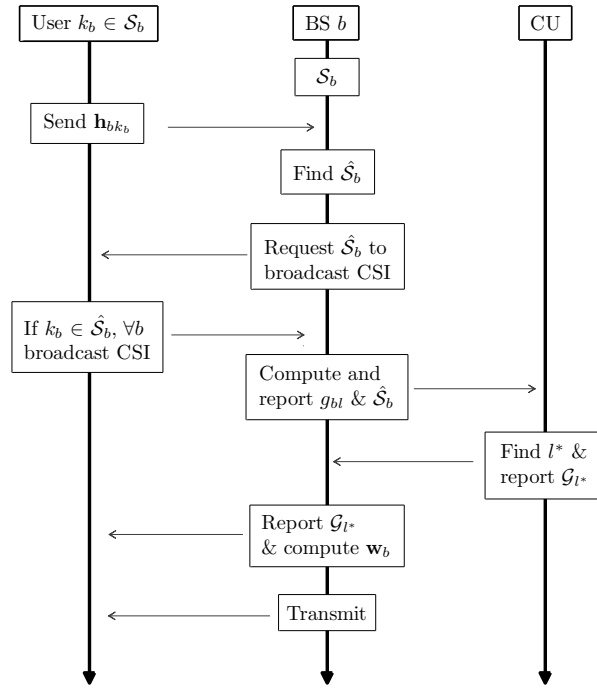


Fig. 1. Proposed Coordinate Scheduling and CBF Transmission

DVSINR the strongest user will be considered for selection. Observe that the index $k_{b(\max)}$ can be one or more of the indices $k_{b(n)} \forall n$ which may be repeated as well, and $|\hat{\mathcal{S}}_b|$ can be at most $N_t + 1$. Assuming that $|\mathcal{S}_b| \geq N_t + 1, \forall b$ the minimum number of reported metrics per BS, denoted as L_r , that will be used to solve (23) is bounded as follows:

$$L_r = \prod_{b=1}^B |\hat{\mathcal{S}}_b| \leq (N_t + 1)^B \leq L = \prod_{b=1}^B |\mathcal{S}_b| \quad (27)$$

and notice that L_r is independent of $|\mathcal{S}_b| \forall b$. Fig. 1 illustrates the sequence of information exchange between users, BSs, and the CU considering channel metrics and search space pruning.

VI. NUMERICAL RESULTS

In this section the performance of the joint distributed linear precoding and user selection is illustrated numerically. The results are obtained using the deployment described in Section II with $B = 3$, cell radius $r = 1000(\text{m})$, and cell-edge cooperation area of radius $r_{\text{coop}} = 300(\text{m})$. For simplicity all BSs have the same number of users K . The long-term channel power gain is

proportional to $1/d_{bk_b}^4$ where d_{bk_b} is the distance between user k_b and BS b given in meters. We assume perfect CSI at each BS, the average sum rate is given in (bps/Hz), and the results are averaged over 10,000 channel realizations. The results are computed by assigning $P_b = P$ for all $b \in \{1, \dots, B\}$ and the same SNR regime at the cell border to all BSs, i.e., $\rho = P/\sigma_n^2$.

The system performance benchmark is given by the optimal solution of problem (4) which is achieved by global CSI at the CU and is referred to as O-GCSI. In order to solve problem (23) two strategies are implemented: 1) considering all L user permutations, and 2) applying the search space pruning with L_r user permutations. For scenarios where $N_t \geq B$, the results obtained for (14) are referred to as O-MUS (metric of user selection) when L is considered, or R-MUS if L_r is used. Similarly, metric (21) is referred to as O-NSPA (NSP approximation) for L and R-NSPA for L_r . If $N_t < B$ the results for (18) are referred to as O-MUS2 and R-MUS2 for L and L_r respectively. In order to highlight how the proposed metrics exploit multiuser diversity we compare their performance w.r.t. a selfish user selection where each BS transmits to its strongest user (maximum channel norm) referred to as Max-SNR.

A. Sum rate vs SNR (ρ)

The average sum rate as function of ρ (dB) for DZF and DVSINR is displayed in Fig. 2 and Fig. 3 respectively. In Fig. 2 for the case $N_t = 3$, $B = 3$, and a target rate of 13(bps/Hz) the O-NSPA requires about 1(dB) extra to achieve the target compared to O-GCSI. For a target ρ of 10(dB) the O-NSPA has a gap about 1(bps/Hz) compared to O-MUS. The simulated scenarios considered $K = 10$ users per BS, O-GCSI, O-MUS, O-NSPA require to evaluate $L = 10^3$ metrics per BS while R-MUS, R-NSPA require $L_r \leq 4^3$ for $N_t = 3$. For $\rho = 10$ (dB) R-MUS and R-NSPA achieve 96% and 91% of the optimal performance O-GCSI, which shows the effectiveness of the search space pruning for CBF systems under DZF precoding. The performance gap between R-MUS and R-NSPA w.r.t. O-GCSI is about 1% and 2.5% respectively for $N_t = 4$ and $\rho = 10$ (dB).

Fig. 3 show results for DVSINR with $B = 3$ and $N_t \in \{2, 3\}$. In the case $N_t \geq B$ the results for O-MUS and O-GCSI are quite closed and the sum rate of O-MUS is sub-optimal in the middle SNR range. The performance gap of O-MUS is less than 3% in the SNR range $\rho \in [-10, 10]$ and such a gap vanishes for other values of ρ . In the case of O-NSPA for $N_t \geq B$, it achieves up to 96% of the rate of O-GCSI in the whole SNR range. For $N_t = 3$ and $\rho = 10$ (dB), the performance gap between R-MUS and R-NSPA w.r.t. O-GCSI is about 3% and 5% respectively.

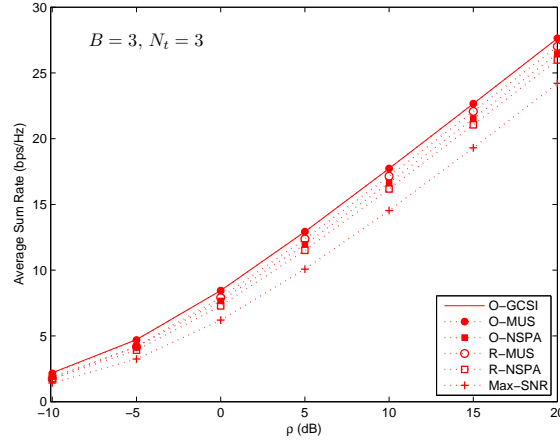


Fig. 2. Average Sum Rate as a function of ρ (dB) for DZF precoding with $K = 10$, $B = 3$ and $N_t = 3$.

We have discussed that for the interference limited scenario ($N_t < B$), the SINR (3) of user $k_b \in \mathcal{S}_b$ depends on all its cross-channels $\{\mathbf{h}_{jk_b}\}_{j=1, j \neq b}^B$. However, BS b only knows \mathbf{h}_{bk_b} and an accurate user selection must take into account both the effective channel gain over the direct and cross channels, unlikely the case $N_t \geq B$. The figure shows that considering all user permutations L for metric (18), O-MUS2, is highly efficient in the low SNR regime and it achieves up to 91% of the sum rate of O-GCSI when $\rho = 20$ (dB). In contrast, O-NSPA cannot exploit multiuser diversity efficiently and only achieves 78% of the O-GCSI performance at the same SNR. Accounting for the search space pruning, R-NSPA and R-MUS2 attain 79% and 73% of the O-GCSI performance, respectively. These results show the effectiveness of the propose metric (18) and highlight the fact that we rely on L metrics per BS in order to achieve acceptable performance and compensate the lack of CSI knowledge of other BSs.

It is worth mentioning that the performance in the interference limited scenario ($N_t < B$) can be improved by joint power allocation (coordinated by the CU) but this requires global CSI knowledge of the scheduled users and their precoder vectors. This kind of power optimization is out of the scope of the paper and we refer to [?] for an in-depth survey.

B. Sum rate vs K

Fig. 4 shows the average sum rate as a function of the number of users (multiuser diversity) K for DZF and $\rho = 10$ (dB). The figure illustrates the average sum rate of two scenarios where

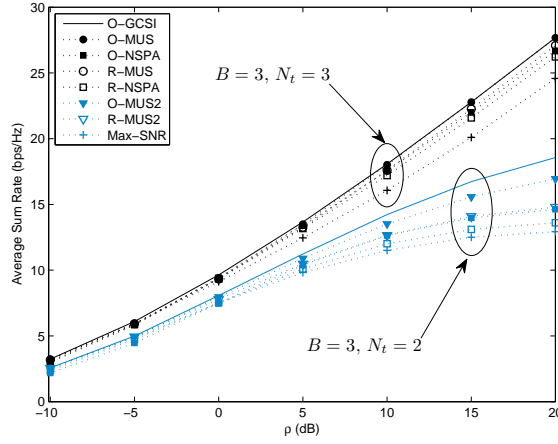


Fig. 3. Average Sum Rate as a function of ρ (dB) for DVSINR precoding with $K = 10$, $B = 3$ and $N_t \in \{2, 3\}$.

$N_t = B$ and $N_t > B$. In our CBF scenario, numerical results show that the set of users maximizing the product of their effective channel gains (computed with local CSI) achieve maximum sum rate for DZF. For $K = 12$ O-MUS and O-GCSI overlap, O-NSPA attains up to 94% of the sum rate of O-GCSI for $N_t = 3$ and 98% for $N_t = 4$. The performance of R-MUS and R-NSPA illustrates the benefits of the proposed search space pruning. For $K = 12$, $N_t = 3$, and $B = 3$ the sum rate gap between R-MUS and O-MUS is less than 4% while the gap between R-NSPA and O-NSPA is less than 3% but the gain in terms of computed metrics per BS is remarkable since $L_r \leq 64 < L = 1728$.

Fig. 5 shows the sum rate as a function of K for DVSINR and $\rho = 10$ (dB). The figure shows the performance for three scenarios with fixed $B = 3$ and $N_t \in \{4, 3, 2\}$. For $N_t = 4$ and $K = 12$, O-MUS and O-NSPA achieve 99% and 98% respectively of the benchmark sum rate. For $N_t = 3$ and $K = 12$, O-MUS and O-NSPA achieve 99% and 97% of the O-GCSI performance respectively. When $N_t = 2$, the O-MUS2 and O-NSPA achieve 94% and 89% of the optimal sum rate, respectively. The computational gains due to search space pruning are larger when $N_t \geq B$ and slightly reduced when $N_t < B$. For the latter scenario the performance gap between O-MUS2 and R-MUS2 is about 7% while the gap between O-NSPA and R-NSPA is about 6% for $K = 12$.

The performance gap between O-MUS and O-NSPA for both DZF in Fig. 4 and DVSINR

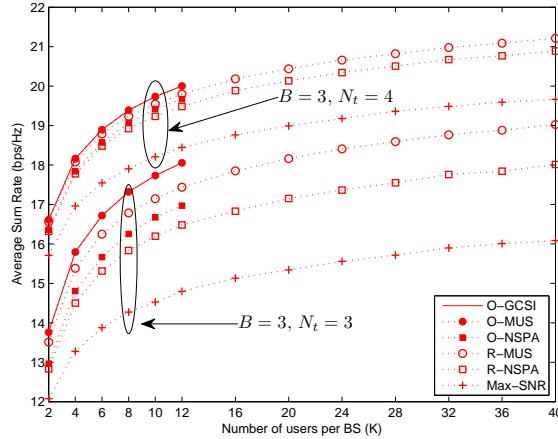


Fig. 4. Average Sum Rate as a function of the number of users per BS (K) for DZF with $\rho = 10(\text{dB})$, $B = 3$, and $N_t \in \{3, 4\}$.

in Fig. 5 reduces considerably by adding one extra antenna per BS. These results suggest that metric (21) should be preferred instead of (14) for system where $N_t \gg B$. The advantage of (21) is that it only requires inner product and vector norm operations while (14) requires several matrix operations. The results illustrate the effectiveness of the proposed metrics, two of them capturing more accurately spatial compatibility of the multiuser channels (14) and (18), and metric (21) that is less computationally demanding and independent of the relation between B and N_t . The advantage of using the search space pruning is explicit if $K > N_t$. For $K = 40$ there exists $L = 40^B$ possible metrics per BS, however, the CU requires at most $(N_t + 1)^B$ metrics per BS in order to perform efficient user selection.

VII. CONCLUSIONS

In this paper we have addressed the sum rate maximization problem for multi-cell systems in the CBF transmission mode with limited message exchange between BSs. Considering that CSI is not globally available we adopt two distributed linear precoding schemes with defined structures, DZF and DVSINR, and discussed the characteristics of their respective effective channel gains. We showed that user selection must be based on the precoding technique that is implemented and channel metrics were designed for two scenarios 1) $N_t \geq B$ and 2) $N_t < B$. The objective of the metrics is to use the local CSI to provide an estimation of the achievable effective channel gains for each precoder technique. Using the metrics at the CU, we designed

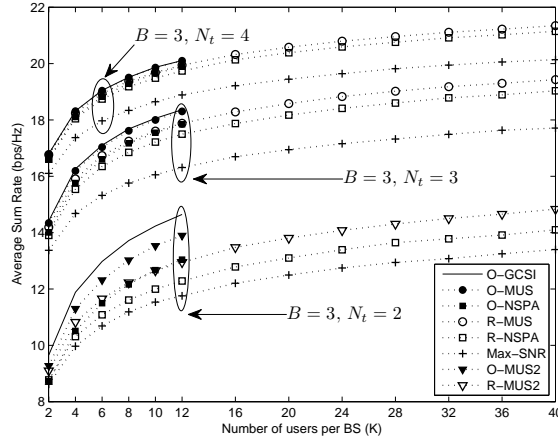


Fig. 5. Average Sum Rate as a function of the number of users per BS (K) for DVSINR with $\rho = 10(\text{dB})$, $B = 3$, and $N_t \in \{2, 3, 4\}$.

an algorithm that selects a set of spatially compatible users. Finally, we proposed a method for search space pruning which dramatically reduces the number of metrics reported from the BSs to the CU and preserves multiuser diversity. Our algorithm and metrics for user selection were assessed by simulations and numerical results show their potential to improve performance in coordinated multi-cell systems with limited message exchange between BSs.

APPENDIX A

Proof of the Proposition 1: In what follows, we slightly abuse the notation and omit the user and BS subindices. Consider the channel of the served user \mathbf{h} , its precoding vector \mathbf{w} defined in (6), and its aggregate interference matrix $\tilde{\mathbf{H}}$, define $\tilde{\mathbf{V}} = \text{null}(\tilde{\mathbf{H}})$ the matrix that contains the orthonormal vectors $\{\tilde{\mathbf{v}}_i\}_{i=1}^{\epsilon}$ and $\rho = P/\sigma_n^2$. The effective channel gain is given by:

$$|\mathbf{h}^H \mathbf{w}|^2 = |\text{Tr}(\mathbf{h} \mathbf{w}^H)|^2 \quad (27a)$$

$$= \frac{|\text{Tr}(\tilde{\mathbf{V}} \tilde{\mathbf{V}}^H \mathbf{h} \mathbf{h}^H)|^2}{|\mathbf{h}^H \tilde{\mathbf{V}} \tilde{\mathbf{V}}^H \mathbf{h}|} \quad (27b)$$

$$= |\mathbf{h}^H \tilde{\mathbf{V}} \tilde{\mathbf{V}}^H \mathbf{h}| \quad (27c)$$

$$= \|\tilde{\mathbf{V}} \tilde{\mathbf{V}}^H \mathbf{h}\|^2 \quad (27d)$$

$$= \left\| \left(\sum_{i=1}^{\epsilon} \tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i^H \right) \mathbf{h} \right\|^2 \quad (27e)$$

$$= \|\mathbf{h}\|^2 \sum_{i=1}^{\epsilon} \cos^2 \theta_{\mathbf{h}\tilde{\mathbf{v}}_i} \quad (27f)$$

where (27a) is due to the fact that $\|\mathbf{w}\| = 1$ and $\langle \mathbf{h}, \mathbf{w} \rangle = \text{Tr}(\mathbf{h}\mathbf{w}^H)$. (27b) is given by substituting \mathbf{w}^H into (27a) and properties of the trace and outer product [26]. (27c) is done by expressing the denominator in (27b) in the form of the numerator. (27d) The basis of the null space projection $\tilde{\mathbf{V}}$ can be used to compute the projection matrix $\mathbf{Q}_h = \tilde{\mathbf{V}}\tilde{\mathbf{V}}^H$ [28, §2.6]. The projector matrix is an idempotent matrix, i.e., $\mathbf{Q}_h = \mathbf{Q}_h\mathbf{Q}_h^H = \mathbf{Q}_h^H\mathbf{Q}_h$ [25]. (27e) is a decomposition of the form $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H = \sum_{i=1}^{\epsilon} \tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i^H$. In (27f), given the orthonormal vectors $\{\tilde{\mathbf{v}}_i\}_{i=1}^{\epsilon}$, the projection of \mathbf{h} onto $\text{Sp}(\tilde{\mathbf{H}})^\perp$ can be computed by the sum of the individual projections onto each one of the orthonormal basis [25]. As $\|\mathbf{h}\|^2$ and $\cos^2 \theta_{\mathbf{h}\tilde{\mathbf{v}}_i}$ are independent variables [29], the expected value of the effective channel gain is

$$\mathbb{E} [\|\mathbf{h}^H \mathbf{w}\|^2] = \mathbb{E} [\|\mathbf{h}\|^2] \mathbb{E} \left[\sum_{i=1}^{\epsilon} \cos^2 \theta_{\mathbf{h}\tilde{\mathbf{v}}_i} \right]. \quad (28)$$

Given $\mathbf{h}, \tilde{\mathbf{v}}_i \in \mathbb{C}^{N_t \times 1}$ define the random variable v_i as

$$v_i = \cos^2 \theta_{\mathbf{h}\tilde{\mathbf{v}}_i}. \quad (29)$$

According to [29] the cumulative probability function of v_i is given by $F_{v_i}(v_i) = 1 - (1 - v_i)^{N_t-1}$ and the expected value of the random variable is $\mathbb{E}[v_i] = \int_0^1 v_i f_{v_i}(v_i) dv_i = \frac{1}{N_t}$.

APPENDIX B

Proof of the Proposition 2:

For the sake of notation, consider the channel of the intended user \mathbf{h} , its aggregate interference matrix $\tilde{\mathbf{H}}$, \mathbf{w} defined in (9), its associated matrix \mathbf{D} , and $\rho = P/\sigma_n^2$. Let $\tilde{\mathbf{H}} = \mathbf{U}_{\tilde{\mathbf{H}}} \mathbf{\Sigma}_{\tilde{\mathbf{H}}} \mathbf{V}_{\tilde{\mathbf{H}}}^H$ be the SVD of the aggregate interference matrix, and the unitary matrix $\mathbf{U}_{\tilde{\mathbf{H}}}$ is formed by the vectors $\{\mathbf{u}_i\}_{i=1}^{N_t}$. The matrix \mathbf{D} can be decomposed as:

$$\mathbf{D} = \sum_{i=1}^{N_t-\epsilon} \lambda_i(\mathbf{D}) \mathbf{u}_i \mathbf{u}_i^H + \sum_{j=N_t-\epsilon+1}^{N_t} \rho \mathbf{u}_j \mathbf{u}_j^H = \mathbf{D}_P + \mathbf{D}_Q \quad (30)$$

The effective channel gain can be expressed as

$$|\mathbf{h}^H \mathbf{w}|^2 = \frac{|\mathbf{h}^H \mathbf{D} \mathbf{h}|^2}{\|\mathbf{D} \mathbf{h}\|^2} = \frac{|\mathbf{h}^H \mathbf{D}_P \mathbf{h} + \mathbf{h}^H \mathbf{D}_Q \mathbf{h}|^2}{\|\mathbf{D}_P \mathbf{h}\|^2 + \|\mathbf{D}_Q \mathbf{h}\|^2} = \|\mathbf{h}\|^2 \beta_{\mathbf{h}} \quad (31)$$

where

$$\beta_{\mathbf{h}} = \frac{\left(\sum_{i=1}^{N_t-\epsilon} \lambda_i(\mathbf{D}) \cos^2 \theta_{\mathbf{h}\mathbf{u}_i} + \rho \sum_{j=N_t-\epsilon+1}^{N_t} \cos^2 \theta_{\mathbf{h}\mathbf{u}_j} \right)^2}{\sum_{i=1}^{N_t-\epsilon} \lambda_i^2(\mathbf{D}) \cos^2 \theta_{\mathbf{h}\mathbf{u}_i} + \rho^2 \sum_{j=N_t-\epsilon+1}^{N_t} \cos^2 \theta_{\mathbf{h}\mathbf{u}_j}}. \quad (32)$$

Observe that all the eigenvalues in (32) are multiplied by squared correlation coefficients. Bearing in mind that $\lambda_i(\mathbf{D})$ and $\cos^2 \theta_{\mathbf{h}\mathbf{u}_i}$ are independent, then approximating the value of $\cos^2 \theta_{\mathbf{h}\mathbf{u}_i}$ by its mean (which is accurate for moderately high values of N_t) we get:

$$\mathbb{E} [\|\mathbf{h}\|^2 \beta_{\mathbf{h}}] \approx \mathbb{E} \left[\|\mathbf{h}\|^2 \frac{\left(\frac{1}{N_t} \sum_{i=1}^{N_t} \lambda_i(\mathbf{D}) \right)^2}{\frac{1}{N_t} \sum_{i=1}^{N_t} \lambda_i^2(\mathbf{D})} \right] \quad (32a)$$

$$= \mathbb{E} [\|\mathbf{h}\|^2 J(\mathbf{eig}(\mathbf{D}))] \quad (32b)$$

where (32a) considers the expected value $\mathbb{E} [\cos^2 \theta_{\mathbf{h}\mathbf{u}_i}] = \frac{1}{N_t} \forall i$ and (32b) follows the definition of the Jain's index in (1).

A numerical example of $\mathbb{E} [|\mathbf{h}^H \mathbf{w}|^2 / \|\mathbf{h}\|^2]$ and its approximation $\mathbb{E} [J(\mathbf{eig}(\mathbf{D}))]$ are presented in Fig. 6 for $B = 3$ and $N_t \in \{3, 4, 6\}$. The curves are normalized regarding to $\|\mathbf{h}\|^2$ in order to exclusively illustrate the relation between the eigenvalues of \mathbf{D} . Notice that the approximation becomes tight as $\rho \rightarrow \infty$ or $\rho \rightarrow 0$, which can be realized by substituting extreme values of ρ in (32) and $J(\mathbf{eig}(\mathbf{D}))$.

APPENDIX C

Proof of the Proposition 3: Let $\tilde{\mathbf{H}} = \mathbf{U}_{\tilde{\mathbf{H}}} \Sigma_{\tilde{\mathbf{H}}} \mathbf{V}_{\tilde{\mathbf{H}}}^H$ be the singular value decomposition (SVD) of the aggregate interference matrix of the served user. And let $\hat{\mathbf{H}} = \tilde{\mathbf{H}} \tilde{\mathbf{H}}^H = \mathbf{U}_{\hat{\mathbf{H}}} \Sigma_{\hat{\mathbf{H}}} \mathbf{V}_{\hat{\mathbf{H}}}^H$ be the SVD of the Hermitian matrix $\hat{\mathbf{H}}$. The diagonal matrix that contains the eigenvalues of $\hat{\mathbf{H}}$ can be defined from the eigenvalues of the aggregate interference matrix as

$$\Sigma_{\hat{\mathbf{H}}} = \Sigma_{\tilde{\mathbf{H}}} \Sigma_{\tilde{\mathbf{H}}}^H, \quad (33)$$

and the eigenvalues of the matrix \mathbf{D} are given by:

$$\lambda_i(\mathbf{D}) = (\rho^{-1} + [\Sigma_{\hat{\mathbf{H}}}]_{ii})^{-1}. \quad (34)$$

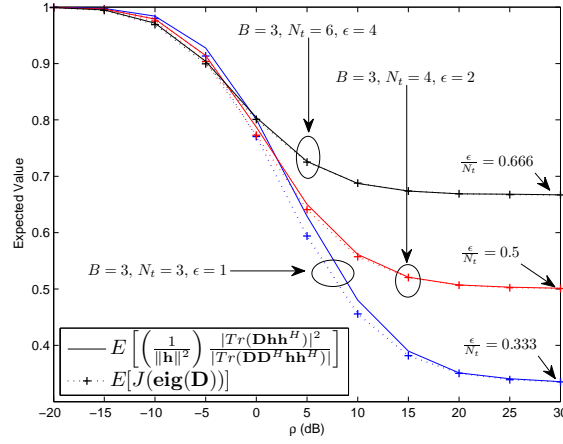


Fig. 6. Normalized values of the effective channel gain of DVSINR precoder and its bound for $B = 3$ and $N_t \in \{3, 4, 6\}$.

Due to the fact that $\lambda_{\min}(\hat{\mathbf{H}})$ is equal to zero with multiplicity ϵ , $\lambda_{\max}(\mathbf{D}) = \rho$ with multiplicity ϵ . This means that $N_t - \epsilon$ eigenvalues of \mathbf{D} are bounded as $\rho \rightarrow \infty$ and ϵ are not. The Jain's index of $\mathbf{eig}(\mathbf{D})$ is such that:

$$\lim_{\rho \rightarrow \infty} J(\mathbf{eig}(\mathbf{D})) = \lim_{\rho \rightarrow \infty} \frac{\left(\sum_{i=1}^{N_t} \lambda_i(\mathbf{D})\right)^2}{N_t \sum_{i=1}^{N_t} \lambda_i^2(\mathbf{D})} \quad (34a)$$

$$= \lim_{\rho \rightarrow \infty} \frac{\left(\sum_{i=1}^{N_t-\epsilon} (\rho^{-1} + [\Sigma_{\hat{\mathbf{H}}}]_{ii})^{-1} + \sum_{j=N_t-\epsilon+1}^{N_t} \rho\right)^2}{N_t \left(\sum_{i=1}^{N_t-\epsilon} (\rho^{-1} + [\Sigma_{\hat{\mathbf{H}}}]_{ii})^{-2} + \sum_{j=N_t-\epsilon+1}^{N_t} \rho^2\right)} \quad (34b)$$

$$= \lim_{\rho \rightarrow \infty} \frac{\left(\sum_{i=1}^{N_t-\epsilon} (\rho^{-1} + [\Sigma_{\hat{\mathbf{H}}}]_{ii})^{-1} + \epsilon \rho\right)^2}{N_t \left(\sum_{i=1}^{N_t-\epsilon} (\rho^{-1} + [\Sigma_{\hat{\mathbf{H}}}]_{ii})^{-2} + \epsilon \rho^2\right)} = \frac{\epsilon}{N_t} \quad (34c)$$

which is illustrated in Fig. 6.

Proof of Proposition 4: In order to simplify the notation let \mathbf{h}_1 be the channel of the user served in the local BS, with its associated matrices $\tilde{\mathbf{H}}_1$ and $\mathbf{D}_1 = \mathbf{D}_{1P} + \mathbf{D}_{1Q}$ as in (30). And let $\mathbf{h}_2 \in \tilde{\mathbf{H}}_1$ be a channel vector used to create the precoding vector \mathbf{w}_1 . The interference term $|\mathbf{h}_2^H \mathbf{w}_1|^2$ in (3) for DVSINR can be unfolded as follows:

$$|\mathbf{h}_2^H \mathbf{w}_1|^2 = \frac{|\mathbf{h}_2^H \mathbf{D}_1 \mathbf{h}_1|^2}{|Tr(\mathbf{D}_1 \mathbf{D}_1^H \mathbf{h}_1 \mathbf{h}_1^H)|} \quad (34d)$$

$$= \frac{\left| \sum_{i=1}^{N_t-\epsilon} \lambda_i(\mathbf{D}_1) \mathbf{h}_2^H \mathbf{u}_i \mathbf{u}_i^H \mathbf{h}_1 \right|^2}{\|\mathbf{h}_1\|^2 \sum_{i=1}^{N_t} \lambda_i^2(\mathbf{D}_1) \cos^2 \theta_{\mathbf{h}_1 \mathbf{u}_i}} \quad (34e)$$

$$= \frac{\left| \sum_{i=1}^{N_t-\epsilon} \lambda_i(\mathbf{D}_1) \langle \mathbf{h}_2, \mathbf{u}_i \rangle \langle \mathbf{u}_i, \mathbf{h}_1 \rangle \right|^2}{\|\mathbf{h}_1\|^2 \sum_{i=1}^{N_t} \lambda_i^2(\mathbf{D}_1) \cos^2 \theta_{\mathbf{h}_1 \mathbf{u}_i}} \quad (34f)$$

$$\leq \|\mathbf{h}_2\|^2 \frac{\left(\sum_{i=1}^{N_t-\epsilon} \lambda_i(\mathbf{D}_1) \cos \theta_{\mathbf{h}_2 \mathbf{u}_i} \cos \theta_{\mathbf{h}_1 \mathbf{u}_i} \right)^2}{\sum_{i=1}^{N_t} \lambda_i^2(\mathbf{D}_1) \cos^2 \theta_{\mathbf{h}_1 \mathbf{u}_i}} \quad (34g)$$

where the numerator in (34e) only takes into account the basis and eigenvalues of \mathbf{D}_{1P} in (30) since \mathbf{D}_{1Q} contains the basis of the null space of \mathbf{h}_2 . The result in (34g) obeys the triangle inequality [26] since the terms $\langle \mathbf{h}_2, \mathbf{u}_i \rangle \langle \mathbf{u}_i, \mathbf{h}_1 \rangle$ in (34f) are complex numbers, and by taking their associated norms and coefficients of correlation their absolute values are already computed, cf. (19).

In order to define an upper bound of $\mathbb{E} [|\mathbf{h}_2^H \mathbf{w}_1|^2]$ notice that the eigenvalues of \mathbf{D}_1 in the denominator of (34g) are affected by an independent random variables of the form (29) with expected value $\frac{1}{N_t}$. We take the upper bound of the numerator of (34g) as follows. The terms $\cos \theta_{\mathbf{h}_1 \mathbf{u}_i}$ and $\cos \theta_{\mathbf{h}_2 \mathbf{u}_i}$ are independent so that $\mathbb{E} [(\cos \theta_{\mathbf{h}_1 \mathbf{u}_i} \cos \theta_{\mathbf{h}_2 \mathbf{u}_i})^2] = \mathbb{E} [\cos^2 \theta_{\mathbf{h}_1 \mathbf{u}_i}] \mathbb{E} [\cos^2 \theta_{\mathbf{h}_2 \mathbf{u}_i}]$. Since $0 \leq \cos^2 \theta_{\mathbf{h}_2 \mathbf{u}_i} \leq 1$ we have that following upper bound $\mathbb{E} [(\cos \theta_{\mathbf{h}_1 \mathbf{u}_i} \cos \theta_{\mathbf{h}_2 \mathbf{u}_i})^2] \leq \mathbb{E} [\cos^2 \theta_{\mathbf{h}_1 \mathbf{u}_i}] = \frac{1}{N_t}$. Considering that the terms $\cos \theta_{\mathbf{h}_1 \mathbf{u}_i} \forall i$ are independent of $\|\mathbf{h}_2\|^2$ and $\text{eig}(\mathbf{D}_1)$, the expected value of the leakage is upper bounded as follows:

$$\mathbb{E} [|\mathbf{h}_2^H \mathbf{w}_1|^2] \leq \mathbb{E} \left[\|\mathbf{h}_2\|^2 \frac{|Tr(\mathbf{D}_{1P})|^2}{Tr(\mathbf{D}_1 \mathbf{D}_1^H)} \right] \quad (35)$$

For the scenarios where $N_t \geq B$ the trace ratio in (35) can be approximated by dividing the largest squared eigenvalue of the numerator given by $\lambda_{\max}^2(\mathbf{D}_{1P}) = (\rho^{-1} + \lambda_{\min}(\tilde{\mathbf{H}}_1^H \tilde{\mathbf{H}}_1))^{-2}$ over the largest squared eigenvalue in the denominator $\lambda_{\max}^2(\mathbf{D}_1) = \rho^2$ which has multiplicity ϵ . By considering only these eigenvalues in the ratio, the contribution of the other eigenvalues is ignored and the approximated expected value of the leakage is given by:

$$\mathbb{E} [|\mathbf{h}_2^H \mathbf{w}_1|^2] \approx \mathbb{E} \left[\frac{\|\mathbf{h}_2\|^2 \lambda_{\max}^2(\mathbf{D}_{1P})}{\epsilon \lambda_{\max}^2(\mathbf{D}_1)} \right] \quad (35a)$$

$$= \mathbb{E} \left[\frac{\|\mathbf{h}_2\|^2}{\epsilon (\rho \lambda_{\min}(\tilde{\mathbf{H}}_1^H \tilde{\mathbf{H}}_1) + 1)^2} \right] \quad (35b)$$

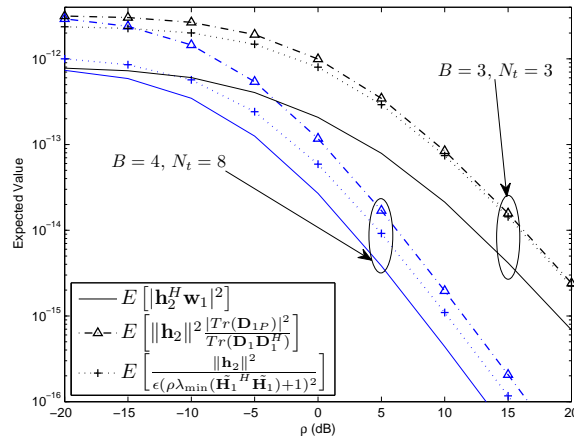


Fig. 7. Upper bound and exact value of the average leakage $\mathbb{E}[|\mathbf{h}_2^H \mathbf{w}_1|^2]$ for $B = 3$ with $N_t = 3$ and $B = 4$ with $N_t = 8$.

A numerical example of $\mathbb{E}[|\mathbf{h}_2^H \mathbf{w}_1|^2]$, its upper bound (35), and the approximation (35b) is presented in Fig. 7.

APPENDIX D

Proof of Proposition 5: From (30) it can be observed that the basis of both $Sp(\tilde{\mathbf{H}})$ and $Sp(\tilde{\mathbf{H}})^\perp$ are combined when forming the precoder. The value of the effective channel gain is a function of ρ and for the low SNR regime \mathbf{D}_P is dominant while in the high SNR regime \mathbf{D}_Q is the dominant term of \mathbf{D} . The exact interaction between the vectors $\{\mathbf{u}_i\}_{i=1}^{N_t}$ and \mathbf{h} is given by $\beta_{\mathbf{h}}$. Observe that the components of \mathbf{D}_Q in (32) compute the magnitude of the projection of \mathbf{h} onto each basis of $Sp(\tilde{\mathbf{H}})^\perp$, the exact NSP component scaled by ρ . The components of \mathbf{D}_P do not represent the exact projection of \mathbf{h} onto $Sp(\tilde{\mathbf{H}})$ because each one of the basis is affected by a different eigenvalue $\lambda_i(\mathbf{D})$. The term $|\mathbf{h}^H \mathbf{w}|^2$ combines a component of \mathbf{h} onto $Sp(\tilde{\mathbf{H}})^\perp$ and weighted components of \mathbf{h} onto the basis of $Sp(\tilde{\mathbf{H}})$.

The intuition behind the heuristic metric (14) is that the effective channel gain is bounded as follows:

$$\|\mathbf{Q}_{\mathbf{h}}\mathbf{h}\|^2 \leq |\mathbf{h}^H \mathbf{w}|^2 \leq \|\mathbf{h}\|^2 \quad (36)$$

This means that one can always take into account the magnitude of $\|\mathbf{Q}_{\mathbf{h}}\mathbf{h}\|^2$ and the component $\|\mathbf{P}_{\mathbf{h}}\mathbf{h}\|^2$ should be modified by a monotonic decreasing function of ρ with values in the range

$[0, 1]$. By observing that $\beta_{\mathbf{h}}$ is the ratio of the squared combination of the eigenvalues of \mathbf{D} over the combination of its squared eigenvalues, the function (15) is defined by the quotient of $\lambda_{\min}^2(\mathbf{D}) = (\rho^{-1} + \lambda_{\max}(\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}))^{-2}$ over $\lambda_{\max}^2(\mathbf{D}) = \rho^2$. The objective of such ratio is to measure how much the maximum and minimum eigenvalues of \mathbf{D} spread out as a function of ρ . Observe that as $\rho \rightarrow 0$ the value of (15) goes to 1 and when $\rho \rightarrow \infty$ the function goes to zero.

Proof of Proposition 6:

Let $\tilde{\mathbf{V}}_{bk_b}(\mathcal{G}_l) = \text{null}(\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l))$ be the matrix that contains the orthonormal basis of the null space of $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$ and $\tilde{\mathbf{v}}_i$ is its i th column vector with $i \in \{1, \dots, \epsilon\}$. The NSP can be computed as $\|\mathbf{h}_{bk_b} \mathbf{Q}_{\mathbf{h}_{bk_b}}\|^2 = \|\mathbf{h}_{bk_b}\|^2 \sum_{i=1}^{\epsilon} \cos^2 \theta_{\mathbf{h}_{bk_b} \tilde{\mathbf{v}}_i}$ (see Appendix A). Recall that $\|\mathbf{h}_{bk_b}\|^2$ and $\cos^2 \theta_{\mathbf{h}_{bk_b} \tilde{\mathbf{v}}_i}$ are independent variables [29] and the factors of the product in the RHS of (22) are independent. Assuming that the components of $\tilde{\mathbf{H}}_{bk_b}(\mathcal{G}_l)$ are i.i.d., we have that:

$$\mathbb{E} \left[\prod_{i_b \neq k_b, i_b \in \mathcal{G}_l} \sin^2 \theta_{\mathbf{h}_{bk_b} \mathbf{h}_{bi_b}} \right] = \left(1 - \frac{1}{N_t} \right)^{(B-1)} \quad (37)$$

and $\frac{\epsilon}{N_t} \leq \left(1 - \frac{1}{N_t} \right)^{(B-1)}$ with equality when $B = 2$ for a given N_t . Notice that equality is asymptotically attained for a fixed value of B when $N_t \rightarrow \infty$.

REFERENCES

- [1] E. Hossain, D. Kim, and V. K. Bhargava, *Cooperative Cellular Wireless Networks*. Cambridge University Press, 2011.
- [2] P. Marsch and G. Fettweis, *Coordinated Multi-Point in Mobile Commun.: From Theory to Practice*. Cambridge University Press, 2011.
- [3] D. Nguyen and T. Le-Ngoc, *Wireless Coordinated Multicell Systems: Architectures and Precoding Designs*, ser. Springer-Briefs in Computer Science. Springer, 2014.
- [4] J. Zhang, R. Chen, J. Andrews, A. Ghosh, and R. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, April 2009.
- [5] S.-H. Park, H. Park, H. Kong, and I. Lee, "New Beamforming Techniques Based on Virtual SINR Maximization for Coordinated Multi-Cell Transmission," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1034–1044, March 2012.
- [6] E. Bjornson and E. Jorswieck, *Optimal Resource Allocation in Coordinated Multi-Cell Systems*, ser. Foundations and Trends(r) in Communications and Information. Now Publishers Incorporated, 2013.
- [7] C. Yang, S. Han, X. Hou, and A. Molisch, "How do we design CoMP to achieve its promised potential?" *IEEE Wireless Commun.*, vol. 20, no. 1, pp. 67–74, February 2013.
- [8] E. Bjornson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative Multicell Precoding: Rate Region Characterization and Distributed Strategies With Instantaneous and Statistical CSI," *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4298–4310, 2010.

- [9] H. Huh, A. Tulino, and G. Caire, "Network MIMO With Linear Zero-Forcing Beamforming: Large System Analysis, Impact of Channel Estimation, and Reduced-Complexity Scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, 2012.
- [10] M. Hong and Z.-Q. Luo, "Academic Press Library in Signal Processing: Communications and Radar Signal Processing," S. Theodoridis and R. Chellappa, Eds. Elsevier Science, 2013, vol. 2, ch. 8. Signal Processing and Optimal Resource Allocation for the Interference Channel.
- [11] B. Niu, V. Wong, and R. Schober, "Downlink Scheduling with Transmission Strategy Selection for Multi-Cell MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 736–747, February 2013.
- [12] A. Silva, R. Holakouei, and A. Gameiro, "Power allocation strategies for distributed precoded multicell based systems," *EURASIP J. Wireless Commun. and Networking*, vol. 2011, no. 1, p. 1, 2011.
- [13] W. Yu, T. Kwon, and C. Shin, "Multicell Coordination via Joint Scheduling, Beamforming, and Power Spectrum Adaptation," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 1–14, July 2013.
- [14] E. Larsson and E. Jorswieck, "Competition Versus Cooperation on the MISO Interference Channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, September 2008.
- [15] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, "Coordinated Beamforming for MISO Interference Channel: Complexity Analysis and Efficient Algorithms," *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 1142–1157, March 2011.
- [16] Z. Tu and R. Blum, "Multiuser diversity for a dirty paper approach," *IEEE Commun. Letters*, vol. 7, no. 8, pp. 370–372, 2003.
- [17] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528 – 541, march 2006.
- [18] P. Chan and R. Cheng, "Capacity Maximization for Zero-Forcing MIMO-OFDMA Downlink Systems with Multiuser Diversity," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 5, pp. 1880–1889, May 2007.
- [19] M. Fuchs, G. Del Galdo, and M. Haardt, "Low-Complexity Space - Time - Frequency Scheduling for MIMO Systems With SDMA," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2775–2784, Sept 2007.
- [20] E. Castañeda, A. Silva, R. Samano-Robles, and A. Gameiro, "Low-complexity User Selection for Rate Maximization in MIMO Broadcast Channels with Downlink Beamforming," *The Scientific World Journal*, vol. 2014, pp. 1–13, 2013.
- [21] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges," *IEEE Commun. Magazine*, vol. 50, no. 2, pp. 148–155, February 2012.
- [22] M. Sadek, A. Tarighat, and A. Sayed, "A Leakage-Based Precoding Scheme for Downlink Multi-User MIMO Channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, 2007.
- [23] R. Jain, W. Hawe, and D. Chiu, "A Quantitative measure of fairness and discrimination for resource allocation in Shared Computer Systems," DEC-TR-301, Tech. Rep., 1984.
- [24] B. Khoshnevis, W. Yu, and Y. Lostanlen, "Two-Stage Channel Quantization for Scheduling and Beamforming in Network MIMO Systems: Feedback Design and Scaling Laws," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2028–2042, October 2013.
- [25] H. Yanai, K. Takeuchi, and Y. Takane, *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*. Springer, 2011.
- [26] J. Gentle, *Matrix Algebra: Theory, Computations, and Applications in Statistics*. Springer, 2007.

- [27] M. Sharif and B. Hassibi, "A Comparison of Time-Sharing, DPC, and Beamforming for MIMO Broadcast Channels With Many Users," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 11–15, Jan 2007.
- [28] G. Golub and C. Van Loan, *Matrix Computations*, ser. Matrix Computations. Johns Hopkins University Press, 1996.
- [29] C. K. Au-Yeung and D. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, 2007.

